Solutions: Homework Set # 5

Problem 1 (There are almost no perfect codes)

(a) We know that if the code is αN-error-correcting code, then its minimum distance must be greater than $2\alpha N$, namely

$$d_{\min} > 2\alpha N.$$

(b) For each pair of codewords $c_i, c_j \in \mathcal{C}$ we can write

$$\begin{aligned} 2\alpha N &< d_{\min} \\ &\leq d(c_i, c_j). \end{aligned}$$

From the definition of c_0, c_1, and c_2 we have

$$d(c_0, c_1) = (u+v)N, \quad d(c_1, c_2) = (u+w)N, \quad \text{and} \quad d(c_0, c_2) = (v+w)N.$$

But we know that $2\alpha N < d(c_i, c_j)$ so we can write

$$(u+v) > 2\alpha, \quad (u+w) > 2\alpha, \quad \text{and} \quad (v+w) > 2\alpha.$$

Summing these three inequities and dividing by two, we have

$$u + v + w > 3\alpha.$$

So if $\alpha > 1/3$, we can deduce $u + v + w > 1$, so $x < 0$, which is impossible. Such a code cannot exist. So the code cannot have three codewords, let alone 2^{NR_C}.

We conclude that, whereas Shannon proved there are plenty of codes for communicating over a binary symmetric channel with $\alpha > 1/3$, there are no perfect codes that can do this. (An αN-error-correcting code can be used for error-free communication over a binary symmetric channel with flipping probability α.)

Problem 2 (Reed-Solomon Codes)

(a) Suppose that the parity check matrix H is an $(n-k) \times n$ matrix and let us assume h_i, $i = 1, \ldots, n$, are the columns of H, namely

$$H = \begin{bmatrix} | & & | \\ h_1 & \cdots & h_n \\ | & & | \end{bmatrix}_{(n-k)\times n}.$$

We assume that every $d-1$ columns of H are linearly independent. This means that if we know

$$a_1 h_{i_1} + \cdots + a_{d-1} h_{i_{d-1}} = \overrightarrow{0},$$

for some numbers a_i then we conclude that $a_1 = \cdots = a_{d-1} = 0$. Here $\overrightarrow{0}$ denotes the all zero vector of length $n-k$ and the indices $i_1, \ldots, i_{d-1}$, take values from the set $\{1, \ldots, n\}$ and non of them are equal.

We know that for a linear code we define the minimum distance of the code as follows

$$\begin{aligned} d_{\min} &\triangleq \min_{c,c' \in \mathcal{C}, c \neq c'} d(c, c') \\ &= \min_{c,c' \in \mathcal{C}, c \neq c'} d(c - c', 0) \\ &= \min_{c \in \mathcal{C}, c \neq 0} h_w(c), \end{aligned}$$

where $d(\cdot, \cdot)$ is the Hamming distance between two codewords and $h_w(\cdot)$ is the Hamming weight of a codeword, the number of non-zero elements of that codeword.

Now assume that the minimum distance of the code $\mathcal{C}$ represented by the parity check matrix H is less than d. We will show that this assumption leads us to a contradiction. By this assumption we know that there exist a non-zero codeword $c \in \mathcal{C}$ where $h_w(c) < d$. Because c is a codeword it should satisfy the equation

$$Hc^T = \overrightarrow{0}. \tag{1}$$

Let us assume $h_w(c) = k < d$ and let $i_1, \ldots, i_k$, denote the indices of non-zero elements of c. So (1) can be written as follows

$$c_{i_1} h_{i_1} + \cdots + c_{i_k} h_{i_k} = \overrightarrow{0}.$$

We know that $k < d$ so by the assumption of the problem the vectors $h_{i_1}, \ldots, h_{i_k}$, are linearly independent. This means that we should have $c_{i_1} = \cdots = c_{i_k} = 0$. Remember that these are the only non-zero elements of c so we can deduce that the codeword c should be zero which is a contradiction (we had assumed that it was a non-zero codeword).

The above argument shows that the minimum distance of the code $\mathcal{C}$ should be at least d.

(b) Without loss of generality and for the simplicity of notation we only consider the first $n-k$ columns but the argument is true for every subset of columns containing $n-k$ elements.

Let us consider the sub-matrix H' of H that contains the first $n-k$ columns of H. So H' is an $(n-k) \times (n-k)$ square matrix of the following form

$$H' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-k} \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_{n-k}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{n-k-1} & \alpha_2^{n-k-1} & \cdots & \alpha_{n-k}^{n-k-1} \end{bmatrix}_{(n-k)\times(n-k)}.$$

In fact H' is a *Vandermonde matrix*. From the linear algebra we know that a square matrix is full rank (all of the rows are independent with each other and all of the columns are independent from each other) if its determinant is non-zero. So to show that the $n-k$ columns of H' are linearly independent we show that the determinant of H' is non-zero. From the hint of the problem we know that

$$\det(H') = \prod_{1 \leq i < j \leq (n-k)} (\alpha_j - \alpha_i),$$

which is not zero if we have $\alpha_i \neq \alpha_j$ for all $i, j, i \neq j$.

The above argument can be applied to other subset of $n-k$ columns of H exactly in the same way. So up to here we have shown that every $n-k$ columns of H are linearly independent.

(c) From part (a) and part (b) we can deduce that the minimum distance of code $\mathcal{C}$ described by matrix H is at least $n-k+1$, namely

$$d_{\min} \geq n-k+1.$$

On the other hand, from the Singleton bound (which is proved in the class) we know that

$$d_{\min} \leq n-k+1,$$

so for the code $\mathcal{C}$ we have $d_{\min} = n-k+1$, which means that $\mathcal{C}$ is a maximum distance separable code.

Problem 3

(a) For the entropy rate of this Markov chain we can write

$$\begin{aligned} H(\mathcal{S}) &= \lim_{n\to\infty} H(S_n|S_{n-1}, \ldots, S_1) \\ &= \lim_{n\to\infty} H(S_n|S_{n-1}) \\ &= H(S_2|S_1), \end{aligned}$$

where $S_i \in \{00, 01, 10\}$.

Now we can expand $H(S_2|S_1)$ and write

$$H(S_2|S_1) = \sum_{s\in\{00,01,10\}} H(S_2|S_1 = s)\mathbb{P}\left[S_1 = s\right],$$

where $\mathbb{P}\left[S_1 = s\right]$ is the stationary distribution of this Markov chain.

Note that in this Markov chain from every state we can go to two other states. So the maximum value of $H(S_2|S_1 = s)$ is at most 1 bit for every state s. Furthermore we note that we can make $H(S_2|S_1 = s) = 1$ for every s by choosing $p = q = r = 1/2$. In this case the entropy rate of the Markov chain is maximized and we have

$$\begin{aligned} H(S_2|S_1) &= \sum_{s\in\{00,01,10\}} H(S_2|S_1 = s)\mathbb{P}\left[S_1 = s\right] \\ &= \sum_{s\in\{00,01,10\}} 1 \cdot \mathbb{P}\left[S_1 = s\right] \\ &= 1 \text{ bit}. \end{aligned}$$

(b) For every state transition of the Markov chain the source outputs two bits, *i.e.*, X_{2i-1} and X_{2i}. Then we feed these bits into finite state machine depicted in Figure 1. We will

show that this FSM compress the sequence as much as possible. For the entropy rate of the source output, the sequence $X_1, \ldots, X_{2n}$, we have

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{2n} H(X_1, \ldots, X_{2n}) &= \lim_{n\to\infty} \frac{1}{2n} H(S_1, \ldots, S_n) \\ &= \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} H(S_1, \ldots, S_n) \\ &= \frac{1}{2} H(\mathcal{S}) \\ &= 1/2 \text{ bit.}\end{aligned}$$

The FSM given in Figure 1 is designed such that it maps $01 \to 0$, $00 \to 1$, and $10 \to 1$, (note to the properties of source $\{X_i\}_{i=1}^{2n}$). So for the entropy rate of output sequence of FSM $\{Y_i\}_{i=1}^{n}$ we have

$$\begin{aligned}H(\mathcal{Y}) &= \lim_{n\to\infty} \frac{1}{n} H(Y_1, \ldots, Y_n) \\ &= \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_{2n}) \\ &= \lim_{n\to\infty} \frac{1}{n} H(S_1, \ldots, S_n) \\ &= H(\mathcal{S}) \\ &= 1 \text{ bit,}\end{aligned}$$

which is the maximum possible entropy rate of sequence of binary bits. So we have show that by using FSM of Figure 1 we can maximally compress the sequence $\{X_i\}_{i=1}^{2n}$ with entropy rate 1/2 to the sequence $\{Y_i\}_{i=1}^{n}$ with entropy rate 1.

(c) Clearly this FSM is uniquely decodable because by having the initial state of FSM and the output string $\{Y_i\}$ we can decode the sequence of input that is fed to the FSM. Note that having the initial state and observing the output at each time is sufficient to find what is the next state of FSM so at each time we can uniquely determine what was the input to the FSM.

(d) From the lecture we know that the condition of being information loss-less is a necessary condition to have a uniquely decodable encoder so this FSM is also information loss-less.

Problem 4

(a) The stationary distribution is $\pi = [p_0, p_1]$, such that $\pi P = \pi$, where

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{bmatrix}.$$

Thus, $\pi = [\frac{p_{1,0}}{p_{0,1}+p_{1,0}}, \frac{p_{0,1}}{p_{0,1}+p_{1,0}}]$

The form of the sequence of states from state 0 returning to state 0 for the first time would be

$0\overbrace{11...1}^{l}0$ for $l = 0, 1, ...$

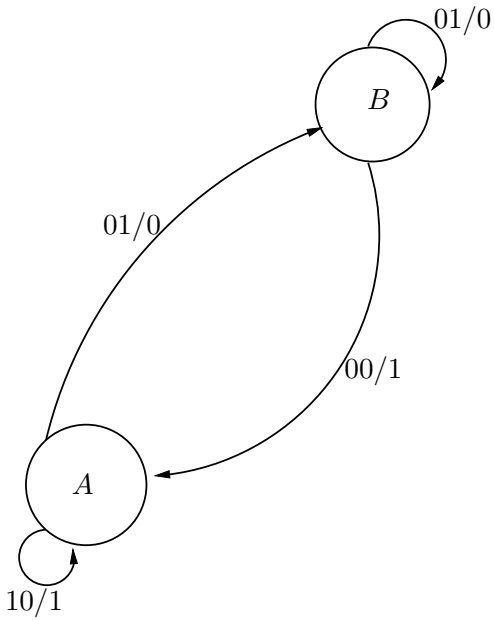

Figure 1: The finite state machine for compressing the source explained in problem 3. Note that the labels (for example, 01/0) on the arrows show the output of the source and the corresponding output of the FSM respectively. For example, if current state $A_i = A$, and the output of the source yields 01, then the FSM outputs 0 and makes a state transition to B, *i.e.,* $A_{i+1} = B$.

And each has a returning time of $l + 1$. So on average we have

$$\mathbb{E}(\text{returning time to } 0) = \sum_l p(X_1...X_{l+2} = 011..10|X_1 = 0) \times (l+1)$$

$$= p(X_1X_2 = 00|X_1 = 0)\times 1 + p(X_1X_2X_3 = 010|X_1 = 0)\times 2 + \sum_{l\geq 2} p(X_1X_2...X_{l+2} = 0\overbrace{11..1}^{l}0) \times (l+1)$$

$$= p_{0,0} + p_{0,1}p_{1,0} \times 2 + \sum_{l=2}^{\infty} (l+1)p_{0,1}(p_{1,1})^{l-1}p_{1,0}$$

$$= p_{0,0} + p_{0,1}p_{1,0}\tfrac{1}{p_{1,1}} \textstyle\sum_{l=1}^{\infty} (l+1)(p_{1,1})^l$$

$$= p_{0,0} + p_{0,1}p_{1,0}\tfrac{1}{p_{1,1}} \underbrace{\sum_{l=1}^{\infty} (p_{1,1})^l}_{=1\frac{p_{1,1}}{1-p_{1,1}}} + p_{0,1}p_{1,0}\tfrac{1}{p_{1,1}} \underbrace{\sum_{l=1}^{\infty} l(p_{1,1})^l}_{=\frac{p_{1,1}}{(1-p_{1,1})^2}(*)}$$

$$= \tfrac{p_{1,0}+p_{0,1}}{p_{1,0}} = \tfrac{1}{p_0}$$

$(*)$: $\sum_{l=1}^{\infty} l(p_{1,1})^l = p_{1,1} + 2 \times p_{1,1}^2 + 3 \times p_{1,1}^3 + ...$

$$= p_{1,1} +$$
$$p_{1,1}^2 + p_{1,1}^2 +$$
$$p_{1,1}^3 + p_{1,1}^3 + p_{1,1}^3 + ...$$
$$\vdots \quad \vdots \quad \vdots + ...$$

$= \frac{p_{1,1}}{1-p_{1,1}} + \frac{p_{1,1}^2}{1-p_{1,1}} + \frac{p_{1,1}^3}{1-p_{1,1}} + ...$

$= \frac{p_{1,1}}{(1-p_{1,1})^2}$

(b) First note that in this expectation it is assumed that $X_0 = 0$. Furthermore, $X_{T_0} = 0$ as well. Thus For $i \neq 0$,

$$s_i = \mathbb{E}_0[\sum_{n\geq 1} 1_{\{X_n=i\}} 1_{\{n\leq T_0\}}] \quad (2)$$

$$= \sum_i \mathbb{E}_0[\sum_{n\geq 0} 1_{\{X_n=i\}} 1_{\{n<T_0\}}] \quad (3)$$

$$(4)$$

So now we prove that $\sum_i s_i p_{i,j} = s_j$

$$\sum_i s_i p_{i,j} = p_{0,j} + \sum_{i\neq 0} \mathbb{E}_0[\sum_{n\geq 1} 1_{\{X_n=i\}} 1_{\{n\leq T_0\}}] p_{i,j} \quad (5)$$

$$= p_{0,j} + \sum_{i\neq 0} \mathbb{E}_0[\sum_{n\geq 0} 1_{\{X_n=i\}} 1_{\{n<T_0\}}] p_{i,j} \quad (6)$$

$$= p_{0,j} + \sum_{i\neq 0} \sum_{T_0>n\geq 0} \mathbb{E} 1_{\{X_n=i\}} p_{i,j} \quad (7)$$

$$= p_{0,j} + \sum_{T_0>n\geq 1} \sum_{i\neq 0} Pr\{X_n = i\} p_{i,j} \quad (8)$$

$$= \sum_{T_0>n\geq 0} \sum_{i\neq 0} [Pr\{X_{n+1} = j\} - Pr\{X_n = 0\} p_{0,j}] \quad (9)$$

$$= \sum_{T_0>n\geq 0} \mathbb{E} 1_{\{X_{n+1}=j\}} \quad (10)$$

$$= \mathbb{E}_0[\sum_{n+1>\geq 1} 1_{\{X_{n+1}=j\}} 1_{\{n+1<T_0\}}] \quad (11)$$

$$= s_j \quad (12)$$

and thus,

$$\sum_i p_i p_{i,j} = \sum_i \frac{s_i p_{i,j}}{\sum_j s_j}$$

$$= \frac{s_j}{\sum_j s_j} = p_j$$

Furthermore, $s_0 = 1$ and $\sum_j S_j = \mathbb{E}(T_0)$ both by definition.

So $p_0 = \frac{1}{\mathbb{E}(T_0)}$ which is the answer to the question, for any general first order Markov process that has a stationary distribution. This is true for any other state as well.

(c) From part (b), we know the average number of the steps of the Markov process to return back to state 0, if we calculate its stationary distribution: (Note that the extended Markov process is formed so that we have a first order Markov process)
$p(x_0^{n-1}) = p(x_0 x_1 ... x_{n-1})$

$= p(x_0) p(x_0 \to x_1) p(x_1 \to x_2) ... p(x_{n-2} \to x_{n-1})$

$p_{x_0} p_{x_0,x_1} p_{x_1,x_2} .... p_{x_{n-2},x_{n-1}}$